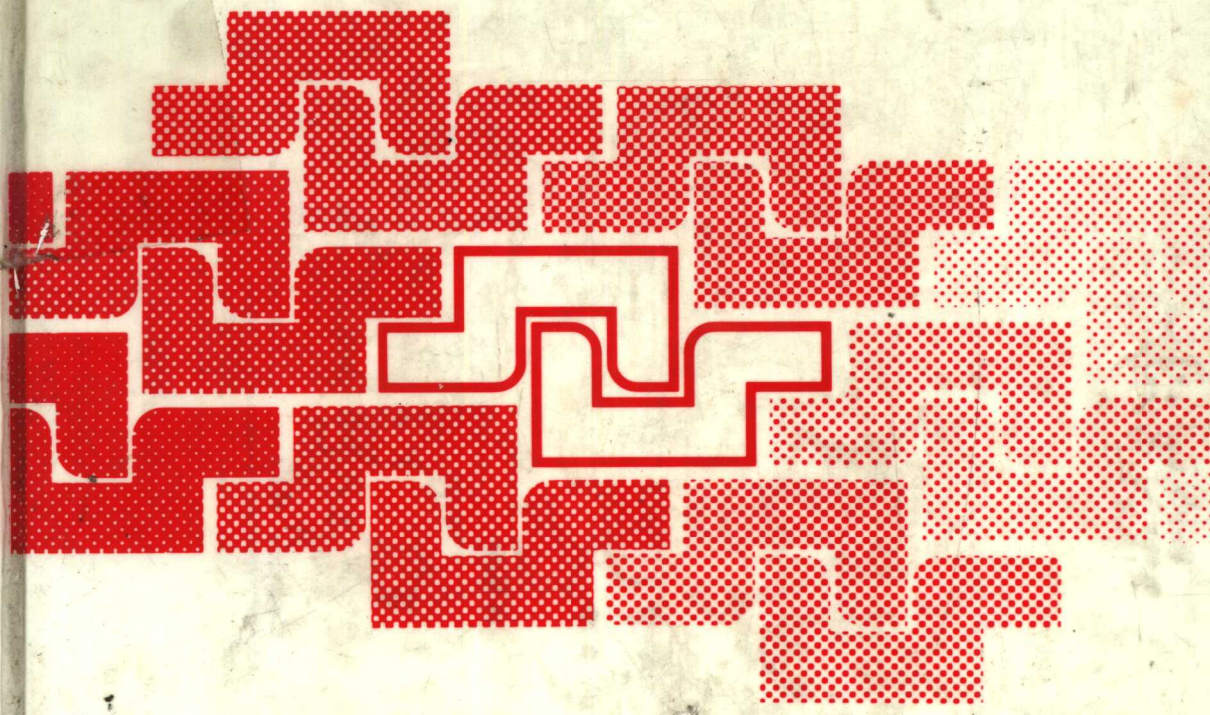


Protein Structure and Engineering

Edited by

Oleg Jardetzky

NATO ASI Series

Series A: Life Sciences Vol. 183

Protein Structure and Engineering

Edited by

Oleg Jardetzky

Stanford University
Stanford, California

Assistant Editor

Robin Holbrook

Stanford University
Stanford, California

Plenum Press
New York and London
Published in cooperation with NATO Scientific Affairs Division

Proceedings of a NATO Advanced Study Institute
and Tenth Course of the International School of
Pure and Applied Biostructure
on Protein Structure and Engineering,
held June 19-30, 1989,
in Erice, Sicily, Italy

Library of Congress Cataloging-in-Publication Data

Protein structure and engineering / edited by Oleg Jardetzky ;
assistant editor, Robin Holbrook.
p. cm. -- (NATO ASI series. Series A, Life sciences ; vol.
183)
"Proceedings of a NATO Advanced Study Institute and Tenth Course
of the International School of Pure and Applied Biostructure on
protein structure and engineering, held June 19-30, 1989, in Erice,
Sicily, Italy"--T.p. verso.
"Published in cooperation with NATO Scientific Affairs Division."
Includes bibliographical references.
ISBN 0-306-43484-9
1. Protein engineering--Congresses. 2. Proteins--Structure-
-Congresses. I. Jardetzky, Oleg. II. Holbrook, Robin.
III. International School of Pure and Applied Biostructure. Course
(10th : 1989 : Erice, Sicly) IV. North Atlantic Treaty
Organization. Scientific Affairs Division. V. Series: NATO ASI
series. Series A, Life sciences ; v. 183.
TP248.65.P76P77 1990
660'.63--dc20 89-49339
CIP

Protein Structure and Engineering

NATO ASI Series

Advanced Science Institutes Series

A series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A	**Life Sciences**	Plenum Publishing Corporation
B	**Physics**	New York and London
C	**Mathematical and Physical Sciences**	Kluwer Academic Publishers Dordrecht, Boston, and London
D	**Behavioral and Social Sciences**	
E	**Applied Sciences**	
F	**Computer and Systems Sciences**	Springer-Verlag
G	**Ecological Sciences**	Berlin, Heidelberg, New York, London,
H	**Cell Biology**	Paris, and Tokyo

Recent Volumes in this Series

Volume 177—Prostanoids and Drugs
edited by B. Samuelsson, F. Berti, G. C. Folco, and G. P. Velo

Volume 178—The Enzyme Catalysis Process: Energetics, Mechanism, and Dynamics
edited by Alan Cooper, Julien L. Houben, and Lisa C. Chien

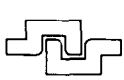
Volume 179—Immunological Adjuvants and Vaccines
edited by Gregory Gregoriadis, Anthony C. Allison, and George Poste

Volume 180—European Neogene Mammal Chronology
edited by Everett H. Lindsay, Volker Fahlbusch, and Pierre Mein

Volume 181—Skin Pharmacology and Toxicology: Recent Advances
edited by Corrado L. Galli, Christopher N. Hensby, and Marina Marinovich

Volume 182—DNA Repair Mechanisms and their Biological Implications in Mammalian Cells
edited by Muriel W. Lambert and Jacques Laval

Volume 183—Protein Structure and Engineering
edited by Oleg Jardetzky



Series A: Life Sciences

ACKNOWLEDGEMENTS

The initiative for the 10th Course of the International School of Pure and Applied Biostructure, "PROTEIN STRUCTURE and ENGINEERING", came from Professor Claudio Nicolini, Director of the International School of Pure and Applied Biostructure, who also took the responsibility for obtaining funding and served as Co-Director of the Course. The course was held at the Ettore Majorana Center for Scientific Culture in Erice, Italy 19 - 30 June 1989 as a NATO Advanced Study Institute. The organizers and participants are also indebted to Professor A. Zichichi, Director of the Ettore Majorana Centre for Scientific Culture, for his support and to the staff of the Centre, notably Dr. Gabriele, Dr. J. Pilarski and Dr. P. Sevelli for their outstanding management of the course administration and all practical arrangements.

CONTENTS

MOLECULAR DYNAMICS

SITE-DIRECTED MUTAGENESIS

INTRODUCTION

The development of molecular biology over the past thirty years has lead to an explosive growth of knowledge of protein structure and of methods for the design and manufacture of proteins of almost any desired sequence. While this seemingly opens limitless possibilities for the engineering of proteins with novel structures and functions, progress is in reality limited by our inability to predict function from structure, or even structure from sequence. Some progress in the understanding of protein folding and of the determinants of biological specificity has been made, but in essence these remain among the major unsolved problems of molecular biology. Therefore, no contemporary discussion of the subject can give the newcomer to the field a recipe for making a protein that will do exactly what he wants it to do. The best one can hope for is to survey the limits of the known in order to chart a few new inroads into the unknown.

The 10th Course of the International School of Pure and Applied Biostructure "PROTEIN STRUCTURE and ENGINEERING", held at the Ettore Majorana Center in Erice, Italy 19 - 30 June 1989 as a NATO Advanced Study Institute, was organized around three sets of issues:

(1) How is one to know what structures to make?
(2) How can one make them?
(3) What potential applications can be expected?

One might note that at this time only a partial answer to the first question can be given; it is possible to describe the methods available for the study of structures and their behavior - these include X-ray diffraction, High Resolution NMR, as well spectroscopic and theoretical methods for the study of molecular dynamics; it is also possible to cite specific examples in which site-directed mutants, designed from prior knowledge of the parent structure, have been used to test various hypotheses about the structure, dynamics, folding or function of a particular protein. For reasons noted above, it is not yet possible to define any kind of general design or architectural principles that would be useful in making decisions for the design of proteins with novel functions at will.

Techniques of genetic engineering have provided many specific answers to the second question. It would be only a slight overstatement to say that, the limitations of existing methods notwithstanding, almost any desired protein sequence could be produced at this time, using existing technology. If the published volume of the proceedings were to accurately reflect the state of knowledge in the field, description of such techniques and their known use would be the by far dominant section. The decision of the organizers was however to focus attention on answering the as yet unanswered questions of rational design and on structural studies that can allow it. For this reason only a sample of the available preparative methods have been included.

Knowledge of protein design and architectural principles will be necessary to transcend the narrow limits within which the third question can now be answered. The use of genetic engineering methods in the manufacture of known proteins for a variety of pharmacological, agricultural and industrial uses is a reality, but hardly involves any protein design. For the present, realistic answers to the question of new applications are limited to modification of known proteins to endow them with modified and, occasionally, new functions. The recent work on catalytic antibodies may serve as a prime example. Several other examples can be found in the present volume. The more bold and speculative suggestions made for stimulating conversation at the meeting, but did not crystallize in a form fit to print.

While the modification of known proteins may seem as a modest framework, compared to the limitless opportunities one can imagine, it must be noted that it has a very important role to play. It can serve to answer the key questions that need to be answered before one could speak of protein architectural principles and embark on a grand scheme of new protein design: What contribution does a particular amino acid at a particular point in the sequence make to the stability of a folded structure - or to the folding pathway? What is its contribution to a specific interaction with a ligand (be it a substrate, inhibitor or regulator)? The number of combinations to be considered in any attempt to answer these and related questions is staggering and, despite the rapidly growing literature, the surface of the problem has barely been scratched.

No single conference and no single volume can aspire to present an encyclopedic review of the large number of current studies relevant to the subject. At best one can hope to present a small selection representative of the cutting edge of the field. It is in this spirit that this volume is offered to the reader.

Oleg Jardetzky
September 29, 1989

EXPLOITATION OF GEOMETRIC REDUNDANCIES AS A SOURCE OF PHASE INFORMATION IN X-RAY STRUCTURE ANALYSIS OF SYMMETRIC PROTEIN ASSEMBLIES - Including a worked example: the three-dimensional structure of the icosahedral β_{60} capsid of heavy riboflavin synthase from *Bacillus subtilis*

Rudolf Ladenstein* and Adelbert Bacher**

*Max-Planck-Institut für Biochemie, D-8033 Martinsried & **Institut für Organische Chemie und Biochemie der Technischen Universität München D-8046 Garching, West Germany

INTRODUCTION

Scope of the Work

The construction of symmetric structures from asymmetric building blocks represents an important feature of nature and is studied by several disciplines of science from different viewpoints. In the field of molecular biology the symmetries of complex macromolecules are of special interest. They constitute the basis for structural organization and biological function in many cases. Maximum stability in oligomeric macromolecules is usually achieved by arranging the subunits in a symmetrical manner such that all of the subunits can form equivalent contacts.

During the past decades of research on biological macromolecules evidence has accummulated that icosahedral symmetry is an important feature which governs the self-organization of protein monomers in the formation of highly symmetric oligomeric complexes. The crystallographic work on virus structures is presently revealing the beauty, complexity and functionality of large macromolecular assemblies (Harrison, 1984; Liljas, 1986). The structure analytic study on heavy riboflavin synthase described in this paper will show that icosahedral symmetry may also be of importance for the structural organization of a bifunctional enzyme complex.

The known symmetries of a complex protein oligomer may very much benefit the determination of its three-dimensional structure. The well-known Patterson search methods (Huber, 1985) represent efficient correlation procedures which enable the crystallographer to extract the symmetry relations among the subunits of a crystalline oligomeric macromolecule from the crystallographic intensity data alone without prior conditions. The knowledge of these symmmetries in turn allows one to use the geometric redundancy in the intensity data set of a crystalline macromolecule in order to derive new structural information by averaging (Bricogne, 1976) of electron density maps in real space. The applicability of these methods has profitted from the development of efficient computers with high storage capacity which have made it possible to treat the vast experimental data in short time and with high accuracy.

The complex structures of highly symmetric protein molecules are fascinating in that they provide a picture of the immense potential for self-organization inherently present in matter.

Furthermore, the detailed knowledge of the structure of a macromolecular system serves as an important basis for the deeper understanding of its function. In most cases it will allow the intelligent planning of investigations with complementary methods, which may provide, together with the structure data, an insight into the complicated structure-function relations of a macromolecular assembly.

Heavy Riboflavin Synthase and Related Macromolecules

Heavy riboflavin synthase (HRS) from *Bacillus subtilis* is a bifunctional enzyme complex with a molecular weight of 10^6 Daltons. It is composed of 60 identical β subunits (M_β = 16200) which form an icosahedral capsid that encloses a trimer of a subunits (M_α = 23500, Bacher et al. 1980). It has been shown by immunochemical methods that the immunological determinants of the α_3 trimers are not accessible for specific antibodies in the native complex $\alpha_3\beta_{60}$ (Bacher et al., 1980; Bacher et al., 1986). On the basis of electron microscopic data (Bacher et al., 1980) and X-ray small angle scattering (Ladenstein et al., 1986) a particle diameter of approximately 150 Å has been derived. The complex $\alpha_3\beta_{60}$ is stable only in a rather narrow pH region around pH 7. Dependent on pH and the concentration of specific substrate- and product analogous ligands (see Fig. 2) disaggregation of the native complex but also reaggregation to stable β_{60} aggregates (26 S), which are characterized by a hollow sphere shape, can occur (Bacher et al., 1986). In the absence of the stabilizing ligands polydisperse mixtures of large β aggregates are formed. The dominating species is characterized by an approximate particle diameter of 290 Å. Its architecture presumably follows the construction principles of truncated icosahedrons (Bacher et al., 1986). The well characterized reactions leading to related β subunit assemblies are shown in Figure 1.

Thus the complex $\alpha_3\beta_{60}$ represents an ideal system for the study of protein-protein and protein-ligand interactions and the self-assembly of macromolecular systems with icosahedral symmetry.

The Catalytic Reaction

The bifunctional complex $\alpha_3\beta_{60}$ catalyzes the final reactions in the biosynthesis of riboflavin (vitamin B_2). Briefly, the β subunits catalyze the condensation of a 3,4 - dihydroxy - butanone 4 - phosphate (**1**) with 5-amino-6-ribitylamino-2,4 (1H,3H) - pyrimidinedione (**2**) yielding 6,7-dimethyl-8-ribityllumazine (**3**) (Bacher et al., 1978; Neuberger et al., 1986). The subsequent dismutation of **3** is catalyzed by the α subunits yielding riboflavin (**4**) and the pyrimidinedione **2** which can be reutilized by the β subunits (Bacher et al.), (Figure 2). The kinetics of the catalytic steps are incompletely understood.

Crystals of Heavy Riboflavin Synthase

The complex $\alpha_3\beta_{60}$ could be crystallized from 1.35 M phosphate buffer pH 8.7 in the presence of 0.5 mM ligand **5** (Ladenstein et al., 1983). The increased pH stability under the influence of the substrate analogous ligand turned out to be crucial for successful crystallization. The crystals diffract X-rays to a resolution of 3.3 Å and belong to space group $P6_322$ of the hexagonal system. The unit cell dimensions are **a** = **b** = 156.4 Å, **c** = 298.5 Å, $\alpha = \beta = 90°$, $\gamma = 120°$. As a consequence of space group symmetry, particle dimensions and threefold particle symmetry, the particle centers must sit on points with symmetry [32]. Thus the crystalline packing may be described either by hexagonal densest packing or by packing in hexagonal layers (Ladenstein et al., 1983). In Figure 3 these two possibilities are shown. By electron microscopic investigation of freeze-etched 3D-crystals the packing in hexagonal layers could be verified (Ladenstein et al., 1986).

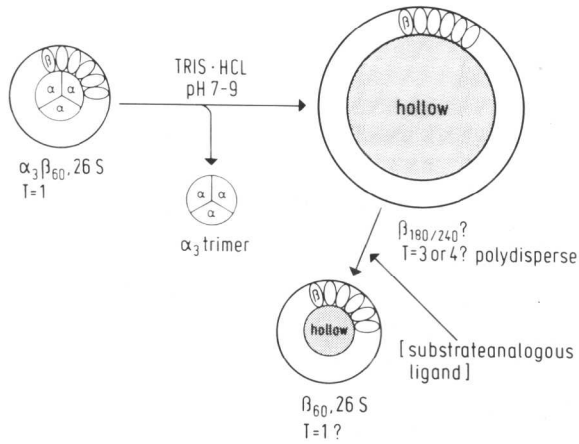


Figure 1. Disaggregation and reaggregation of heavy riboflavin synthase ($\alpha_3\beta_{60}$); formation of hollow β_{60} (26 S) particles.

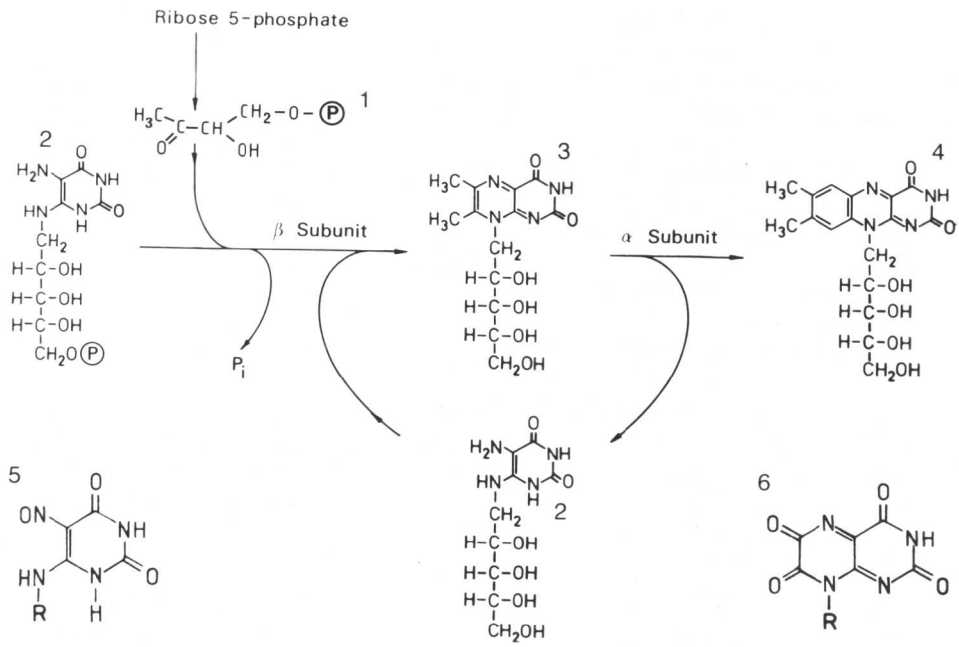


Figure 2. Biosynthesis of riboflavin: **1** = 3,4 - dihydroxy - 2 - butanone 4 - phosphate; **2** = 5-amino-6-(D-ribitylamino)-2,4(1H,3H)-pyrimidinedione; **3** = 6,7-dimethyl-8-(-ribityl)-lumazine; **4** = riboflavin.
Inhibitors of heavy riboflavin synthase: **5** = 5-nitroso-6-ribitylamino-2,4(1H,3H)-pyrimidinedione; **6** = 6,7-dioxo-8-ribityl-5,6,7,8-tetrahydrolumazine.

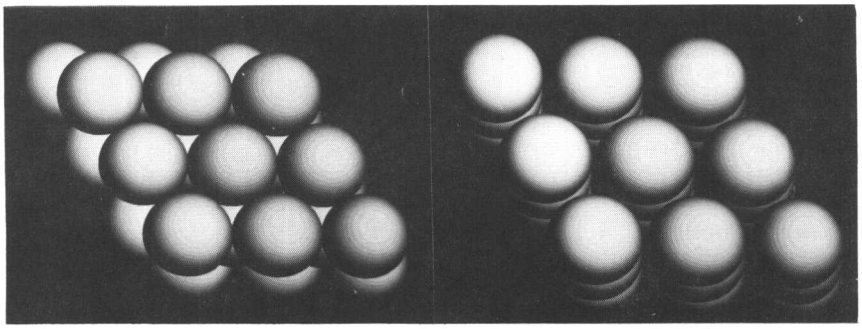


Figure 3. Sphere models showing possible crystalline packing of $\alpha_3\beta_{60}$ particles in the hexagonal unit cell. a) densest hexagonal packing; b) layer packing.

CRYSTALLOGRAPHIC FUNDAMENTALS

Diffraction of X-Rays by a Periodic Object

When a plane wave is scattered by an object, the scattered radiation may be described by the equation

$$\mathbf{F(h)} = \int \rho(\mathbf{x}) \cdot \exp 2\Pi i \mathbf{hx}\, d\mathbf{x} \qquad [1]$$

where **F** is a complex number which represents the amplitude and phase of the scattered radiation in a direction determined by the vector **h**, $\rho(\mathbf{x})$ is the scattering function at a position **x** in the object, and the integral is taken over the volume of the 3D object. For X-ray diffraction, $\rho(\mathbf{x})$ is the electron density at position **x** (excluding the effects of anomalous scattering). Thus the scattered radiation is described by the Fourier transform of the object as seen by the incident radiation. By taking the inverse Fourier transform, we get

$$\rho(\mathbf{x}) = \int \mathbf{F(h)} \cdot \exp -2\Pi i \mathbf{hx}\, d\mathbf{h} \qquad [2]$$

The integral is taken over the volume V* of the space spanned by the vector **h**.

We are interested in the special case in which the scattering object is a 3D crystal. The fundamental property of a crystal is that $\rho(\mathbf{x})$ is periodic in all three dimensions of space. It is known that the Fourier transform of a periodic function is zero, except when **h** is an integer multiple of the periodicity. Thus the structure factors **F(h)** are zero except on a three-dimensional lattice, the so-called reciprocal lattice.

The natural coordinate system for a crystal is

$$\mathbf{x} = x \cdot \mathbf{a} + y \cdot \mathbf{b} + z \cdot \mathbf{c} \qquad [3]$$

where **a,b,c** represent the basis vectors of the unit cell of the crystal. These vectors are not necessarily orthogonal but they define a three-dimensional space, which is called **direct space**. The dimension of direct space is length^3 (L^3). In turn we may define the space of the structure factor lattice, which is related to the recorded diffraction pattern, by

$$\mathbf{h} = h \cdot \mathbf{a}^* + k \cdot \mathbf{b}^* + l \cdot \mathbf{c}^* \qquad [4]$$

with the lengths of the basis vectors **a*,b*,c*** inversely proportional to the lengths of the basis vectors **a,b,c** of the unit cell. The space in which the structure factors are defined is generally called **reciprocal space**; its dimension is length^{-3} (L^{-3}).

In evaluation of diffraction experiments it is often necessary to transform a function defined in reciprocal space into direct space and vice versa. As we have seen these operations can be performed by Fourier transformation (F) and inverse Fourier transformation (F^{-1}) as

$$F[\rho(\mathbf{x})] = \mathbf{F(h)} \qquad [5]$$
$$F^{-1}[\mathbf{F(h)}] = \rho(\mathbf{x}).$$

The Electron Density Function

If we choose **a***, **b*** and **c*** to obey the Laue relations

$$\begin{array}{lll} \mathbf{a}\cdot\mathbf{a}^* = 1 & \mathbf{a}\cdot\mathbf{b}^* = 0 & \mathbf{a}\cdot\mathbf{c}^* = 0 \\ \mathbf{b}\cdot\mathbf{a}^* = 0 & \mathbf{b}\cdot\mathbf{b}^* = 1 & \mathbf{c}\cdot\mathbf{c}^* = 0 \\ \mathbf{c}\cdot\mathbf{a}^* = 0 & \mathbf{c}\cdot\mathbf{b}^* = 0 & \mathbf{c}\cdot\mathbf{c}^* = 1 \end{array} \qquad [6]$$

the vector products in the integrals (1) and (2) simplify to $\mathbf{h}\cdot\mathbf{x} = hx + ky + lz$. With these definitions we are able to normalize (1) and (2) to reflect the contents of one unit cell; we get

$$\mathbf{F(h)} = V \iiint \rho(\mathbf{x})\cdot\exp[2\Pi i(hx + ky + lz)]dxdydz \qquad [7]$$
$$\rho(\mathbf{x}) = 1/V \iiint \mathbf{F(h)}\cdot\exp[-2\Pi i(hx + ky + lz)]dhdkdl \qquad [8]$$

The discrete nature of **F(h)** allows the conversion from an integral to a sum in Eq. [8]

$$\rho(\mathbf{x}) = 1/V \sum_h \sum_k \sum_l \mathbf{F(h)}\cdot\exp[-2\Pi i(hx + ky + lz)] \qquad [9]$$

Equation [9] represents the well-known electron density equation which can be calculated by inverse Fourier transformation of the scattered waves, described by the structure factors **F(h)**. The structure factor **F(h)** is characterized by an amplitude $|\mathbf{F(h)}|$ and a phase $\alpha(\mathbf{h})$ according to

$$\mathbf{F(h)} = |\mathbf{F(h)}|\cdot\exp i\alpha(\mathbf{h}) \qquad [10]$$

The phase information $\alpha(\mathbf{h})$ is lost in the diffraction experiment; phases are generally determined by methods such as single (SIR) and multiple (MIR) isomorphous replacement and phase extension. The amplitudes $|\mathbf{F(h)}| = \text{const.}\cdot\sqrt{\mathbf{I(h)}}$ are obtained from measurement of the crystallographic intensities in diffraction patterns (e.g. photographic rotation method (Arndt et al., 1977), area detectors (Messerschmidt et al., 1987) and diffractometers (Blundell et al., 1976). The set of all symmetry independent **I(h)** represents the unique intensity data set of a crystal. Crystallographic intensity data of native and derivative crystals of heavy riboflavin synthase are shown in Table 1.

Formal Description of Symmetries

By definition the crystallographic symmetries represent the set of symmetry elements valid in a crystal; they relate the asymmetric units of the crystal cell. The noncrystallographic or local symmetry elements are confined to the asymmetric unit and can be described by the set of symmetry operations which are valid in the asymmetric unit. In the case of an oligomeric protein the asymmetric unit of a crystal cell may contain more than one copy of a subunit. The positions of these subunits in direct space are defined by the symmetry operations of the assymetric unit.

Table 1. Intensity data statistics ($F^2 > 1.0\ \sigma$).

Derivative	Resolution	Measurements	Independent Reflections	Measured/Possible Reflections	R_{merge}[%]
ITAN	∞-3.3	75100	27700	0.850 to 3.3Å	13.1
AuCN	∞-3.2	87250	32200	0.673 to 3.2Å	12.8
CMAA	∞-3.6	71300	24800	0.767 to 3.6Å	13.9
WAC	∞-3.4	38200	23000	0.741 to 3.4Å	12.8
WP	∞-3.6	19000	13700	0.434 to 3.6Å	9.4
LUMO	∞-3.6	32500	18300	0.684 to 3.6Å	11.0

ITAN, native crystals; $R_{merge} = \Sigma\Sigma|<I_h>-I_{hi}|/\Sigma N_h<I_h>$, where $<I_h>$ is the average intensity of N_h measurements and I_{hi} is the individual intensity of a reflection h.
LUMO, functional derivative obtained by soaking the native crystals with 1mM of the dioxolumazine (Ligand **6**) in 1.75 M potassium phosphate buffer, pH = 8.7, at 20°C.

Generally we may define a symmetry operation in 3D space by way of a linear transformation including a 3x3 Matrix **R** and a translation vector **t** as

$$\mathbf{x}' = \mathbf{R}\cdot\mathbf{x} + \mathbf{t} \qquad [12]$$

and

$$\rho(\mathbf{x}') = \rho(\mathbf{x}) \qquad [13]$$

for all positions **x** within a crystal or its asymmetric unit.

The Symmetry of Icosahedral Polyhedrons

The ancient Greek mathematicians already knew that only five regular polyhedrons can exist, the so-called platonic solids. These are tetrahedron, octahedron, icosahedron, cube and dodecahedron. Nowadays it is well established that geometry and symmetry of these bodies represent an important principle which governs the self-assembly of protein subunits in quite a large number of cases. A regular icosahedron (Figure 4) is constructed from 20 equilateral triangles and possesses 6 fivefold (n = 72°), 10 threefold (n = 120°) and 15 twofold (n = 180°) rotation symmetry axes. All of these axes intersect at a common point which represents the center (origin) of the particle.

A regular icosahedron is characterized by 60 asymmetric units. This number represents the maximum number of identical units which may be arranged on a closed symmetrical shell. Thus

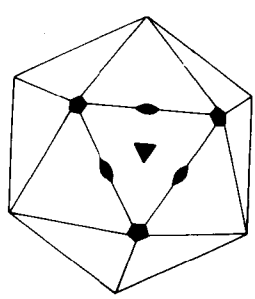


Figure 4. Icosahedron with all symmetry elements of one triangular face indicated: (●) twofold axis, (▼) threefold axis, (⬟) fivefold axis. All of the symmetry axes intersect at the center of the particle: an icosahedral asymmetric unit is defined as the triangular region in between two fivefold and one threefold axis.